

# A Novel Credibility Quantification Method for Welch's Periodogram Analysis Result in Model Validation

Yuchen Zhou, Ke Fang, Kaibin Zhao, Ping Ma\*

Control and Simulation Center, Harbin Institute of Technology, Harbin, P.R. China  
 ZhouYuChen-01@163.com, FangKe@hit.edu.cn, ZhaoKaiBin1986@163.com,  
 PingMa@hit.edu.cn

## Abstract

Welch's periodogram is widely used in frequency domain model validation. However, Welch's analysis results just reveals whether the time series passed the consistency test in each discrete frequency point, which is not a quantitative credibility evaluation result and may not help the evaluation expert to grade credibility level of simulation system. Based on Welch's periodogram and consistency test approach, a novel credibility quantification method using weight density function is proposed. Furthermore, the frequency analysis and credibility quantification process is provided. Finally, the credibility quantification of radiated noise in ship acoustic feature simulation indicates the method proposed is effective for periodic time series with complicated spectrum.

*Keywords:* Welch's periodogram, credibility quantification, line spectrum, periodic time series, model validation

## 1 Introduction

The M&S technology has the advantages of economy, security, repeatability and nondestructive, which makes it widely used in aerospace, nuclear, communication, et al. Verification Validation and Accreditation (VV&A) should be conducted to guarantee the validity of complex simulation system (Oberkampf 2008; Sargent, 2013; Wang, 2000). Through behavior-similarity analysis between simulation system and real-world system, we can obtain the credibility of the simulation system.

Frequency analysis method is usually used for the validation of periodic time series. Fishman and Kiviat firstly proposed frequency domain validation method and applied it to queuing model validation (Fishman, 1967). Gallantetal put forward a data consistency analysis approach based on Analysis of Variance (ANOVA) and periodogram method. Montgomery used spectral analysis to evaluate the credibility of missile simulation system (Montgomery, 1980&1983).

To resolve the thermal challenge problem suggested in reference (Roy, 2011), (Ferson, 2008) proposed an

area metric, which takes the integral over the area difference between the cumulative distribution function(CDF) of simulation data and the empirical CDF of the measured samples as the disagreement between the simulation model and real-world system (Li, 2014; Sankararaman, 2011). Mullins classified the data scenarios with aleatory and epistemic uncertainty and studied how different validation metrics may be appropriate for varies data samples (Mullins, 2015). Literature (Zhang, 2011) provided a group AHP method to evaluate the credibility of complex simulation system, in which Hadamard convex combination is used to aggregate the judgement matrices constructed by different assessment experts.

In the frequency analysis, some spectrum of simulated data and observed data are extremely complicated. Consistency test result of spectrum just shows whether the data passed the examination in each discrete frequency point rather than a quantitative credibility evaluation result, which may not help the evaluation expert to grade the credibility level of simulation system. How to transform the consistency test result to credibility is the key problem to resolve in this paper. The credibility quantitative method is illustrated in detail and case study demonstrates the frequency analysis and credibility transform process of radiated noise comprehensively.

## 2 Periodogram Method

The frequency domain analysis method involves power spectrum density estimation and consistency test. Power spectral density estimation is a data transform approach to estimate the distribution of signal energy in each frequency points using limited data. Welch's periodogram is a typical method in frequency domain validation.

### 2.1 Welch's Periodogram Method

Periodogram method uses Fast Fourier Transform (FFT) algorithm to estimate spectrum of stationary random sequence and the estimated spectrum is sensitive to the length of time series. If the data length is beyond a threshold value, spectrum oscillated

intensely. On the contrary, if the data length is reduced to a certain extent, the resolution of spectral may decrease and the estimation error increase significantly.

Welch's spectral estimation is an improvement of periodogram by dividing the time series into many data segments and using non-rectangular windows to handle each data segments (Welch, 1967). When using Welch's method, the time series  $x(n) \in R^N$  is divided

into  $K = \frac{N}{L}$  pieces, as:

$$x^{(i)}(n) = x(n+iL-L), 0 \leq n \leq L-1, 1 \leq i \leq K \quad (1)$$

Each data segments is including  $L$  samples, and we can calculate  $K$  modified periodogram by:

$$J_L^{(i)} = \frac{1}{LU} \left| \sum_{n=0}^{L-1} x^{(i)}(n)w(n)e^{-j\omega n} \right|^2, i = 1, 2, \dots, K \quad (2)$$

where  $U = \frac{1}{L} \sum_{n=0}^{L-1} w^2(n)$  denotes the mean power of window function.

Finally the power spectrum of time series  $x(n)$  can be calculated by:

$$S_x^w(\omega) = \frac{1}{K} \sum_{i=1}^K J_L^{(i)}(\omega) \quad (3)$$

### 2.2 Consistency Test

Suppose that  $S_x(\omega)$  and  $S_y(\omega)$  are spectrum of simulation model output time series  $x(n)$  and real-world system output time series  $y(n)$ . The estimated spectrum are  $S_x(\omega)$  and  $S_y(\omega)$ . As is proved

$\frac{rS(\omega)}{S(\omega)} \sim \chi_r^2$  (Chen, 1988), we can make the null

hypothesis and alternative hypothesis as:

$$H_0 : S_x(\omega) = S_y(\omega)$$

$$H_1 : S_x(\omega) \neq S_y(\omega)$$

Statistics of hypothesis test is:

$$F = \frac{rS_x(\omega) / S_x(\omega) / r}{rS_y(\omega) / S_y(\omega) / r} \sim F(r, r) \quad (4)$$

where  $r = 2N / \sum_{k=-M}^M w^2(k)$ .

If the original hypothesis is accepted under the confidence level  $\alpha$ , we can draw:

$$p\{F_{\frac{\alpha}{2}, r, r} \leq \frac{S_x(\omega)}{S_y(\omega)} \leq F_{1-\frac{\alpha}{2}, r, r} | H_0\} = 1 - \alpha \quad (5)$$

On the contrary, if  $\frac{S_x(\omega)}{S_y(\omega)} \notin \left[ F_{\frac{\alpha}{2}, r, r}, F_{1-\frac{\alpha}{2}, r, r} \right]$ , we

should accept the alternative hypothesis. Finally, the F test can be made in every frequency point  $\omega_i$ .

## 3 Credibility Quantification Method

### 3.1 Credibility Quantification Method

Since the complexity of computational systems, we may not take the validation result under single input condition as the final credibility of the simulation system. Literature (Mullins, 2015) provides an approach to integrate the model validation results from multiple simulation scenarios, which is defined by equation (6).

$$v_{overall} = \int v(x)\pi(x)dx \quad (6)$$

where  $x$  is a n-dimensional vector of input conditions and  $v(x)$  is the validation result under input condition  $x$ ;  $\pi(x)$  is the joint probability density of the point  $x$ . Mullins et al. propose that  $\pi(x)$  is a weighting function for the importance of multiple validation results which can be estimated according to the relevance of each experiment conditions to the overall intended use of the computational model.

Provided that limited data samples are obtained,  $v(x)$  is available at finite points. Therefore, equation (6) is reformulated as

$$v_{overall} = \frac{\sum_{i=1}^m v(x)w_i}{\sum_{i=1}^m w_i} \quad (7)$$

Literature (Zhang, 2010) studied transform algorithm of several commonly used validation methods. A direct transform approach for periodogram is provided, which takes the percentage of frequency range passed consistency check accounting for the whole frequency range as credibility. This conversion method is based on the hypothesis that statistic characteristic consistency of power spectrum in each frequency point has same effect on the overall credibility. Actually, since the power spectrum density is higher in one or several frequency bands, the effect on the similarity of the consistency test in each frequency points is not equal. Therefore, an inaccurate result may be calculated using direct transform approach.

Power spectrum density curves can be divided into three categories, broadband spectrum (Figure 1), line spectrum (Figure 2), and mixed spectrum (Figure 3) which is composed of broadband spectrum and several line spectrum. In general, the frequency bands with higher energy often reflect the periodic characteristics of the system.

The power difference of line spectrum between predicted time series and observed time series is usually the main reason for failing to pass the consistency check. Thus, the transform of line spectrum should be given high priority in the process of credibility quantification.

Based on the above analysis, the credibility quantification process should focus on the key frequency band with higher power. Therefore, we may use the normal weight density function to realize the

conversion. The integral of normal weight density function  $f(\omega)$  among the whole frequency band is 1 and the multiple of weight function  $f(\omega)$  and consistency test result is the credibility.

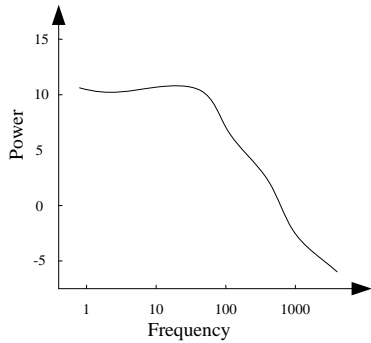


Figure 1. Broadband spectrum.

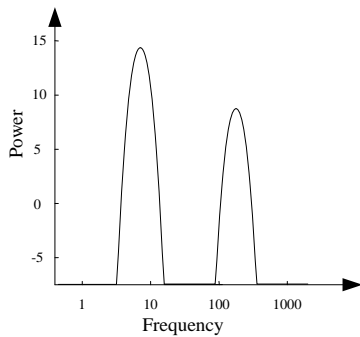


Figure 2. Line spectrum.

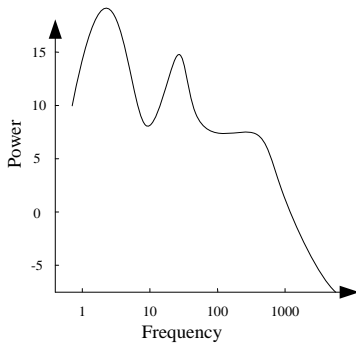


Figure 3. Mixed spectrum,

The normal weight density function is defined as:

$$f(\omega) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\omega - \mu)^2}{2\sigma^2}\right), \omega \in [0, +\infty) \quad (8)$$

where  $\mu$  is the center position of weight density function and  $\sigma$  is the dispersion parameter.

Let  $F(\omega_0)$  be the weight integration of  $f(\omega)$  in  $[0, \omega_0]$ , as:

$$F(\omega_0) = \int_0^{\omega_0} f(\omega) d\omega \quad (9)$$

then the weight integration of  $f(\omega)$  among  $q = [a, b]$  can be expressed as  $F(b) - F(a)$ .

Suppose the frequency points passed the consistency test as 1, otherwise as 0. The whole frequency band is divided into two sections, the band  $M$  passed the check and the band  $N$  failed to pass the check. The two sections are defined as:

$$M = \bigcup_{k=1}^p m_i = \bigcup_{k=1}^p [m_{ka}, m_{kb}] \quad (10)$$

$$N = \bigcup_{k=1}^q n_i = \bigcup_{k=1}^q [n_{ka}, n_{kb}] \quad (11)$$

Then choose appropriate parameters for each weight functions and the integral of band  $M$  is the final credibility.

$$C = F(M) = \sum_{k=1}^p F(m_k) \quad (12)$$

Finally, the parameters of normal weight density function can be chosen as follows.

1) The mean of the normal distribution should be the middle of the line spectrum.

2)  $3\sigma$  Rule may be followed in the selection of the variance.

According to the parameter selection approach, the line spectrum will cover the frequency interval  $[\mu - 3\sigma, \mu + 3\sigma]$ . Although  $\lim_{\omega \rightarrow +\infty} F(\omega) = \int_0^{\omega} f(\omega) d\omega = 1 - \Delta < 1$ , the weight loss  $\Delta < 0.15\%$  and it may be ignored. For example, actuator works periodically in control system. Figure 4 is the power spectrum of actuator output time series. The work frequency of the actuator is between 8 Hz and 10 Hz. Thus, we may choose  $\mu = 9$ ,  $\sigma = 0.67$  as the parameters of weight function (Figure 5).

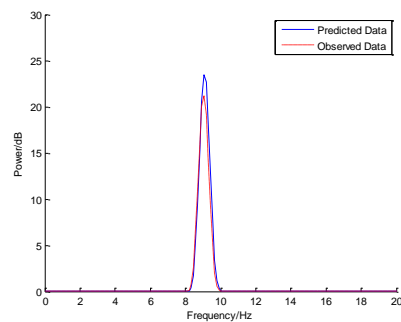


Figure 4. Spectrum of actuator output time series.

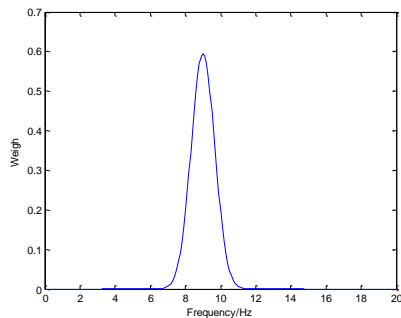


Figure 5. Normal weight density function.

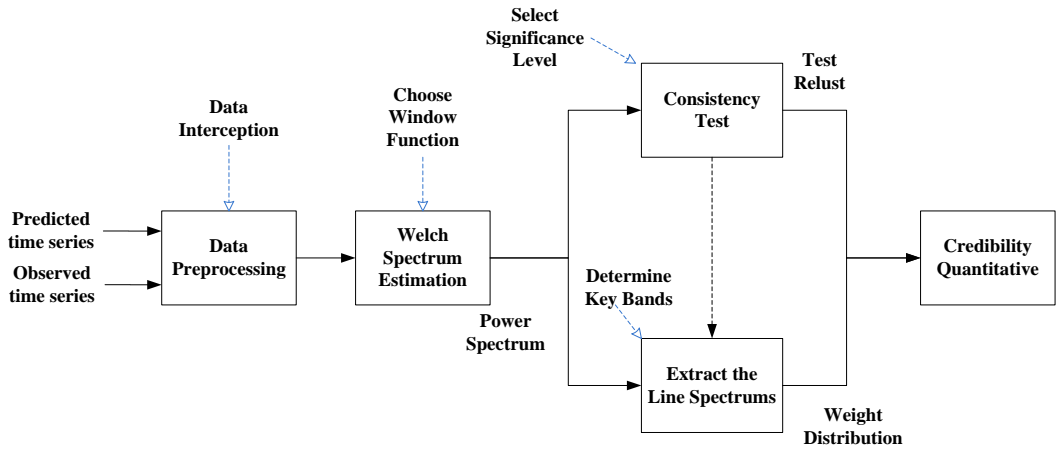


Figure 6. Frequency analysis and credibility quantification process.

### 3.2 Frequency Analysis and Credibility Quantification Process

Based on the above method, including welch analysis method and transform approach, the specific transform steps (Figure 6) are provided as follows.

Step 1. Preprocessing the predicted time series and observed time series, including data interception and smoothing filtering.

Step 2. Choose appropriate transformation points and estimate the frequency spectrum of predicted time series and observed time series.

Step 3. Select a significance level and make the consistency test.

Step 4. Extract line spectrums in power spectrum, and analysis the reasons for the bands failed to pass the consistency test. Based on the test result, separate the whole band into two sections (band *M* passed the consistency test and band *N* failed to pass the consistency test).

Step 5. Select appropriate weight functions for each line spectrums and calculate the final credibility using equation (12).

Actually, the overall simulation credibility of computational model should integrals multiple validation results of simulation data and measured samples under varies input conditions. This paper focus on the credibility quantification of frequency-domain validation result under single data scenario.

### 4 Case study

The task of ship acoustic feature simulation is not only to study and analyze the acoustic characteristics of the ship, but also to realistically simulate the characteristics of the ship and applied it to the sonar system testing and ship type identification. In the research of ship acoustic feature simulation, to guarantee the characteristics similarity of radiated noise between reconstruction model and real ship, frequency domain validation can be conducted to

evaluate the credibility of the ship radiated noise reconstruction model.

Radiated noise is mainly composed of mechanical noise, propeller noise and hydrodynamic noise. In general, radiated noise power spectrum is a typical representative of mixed spectrum. In the radiated noise spectrum, line spectrum reflects the periodic part of the energy distribution of the noise in the signal, which primary covers low frequency band. The line spectrum is the main feature to recognize ship types.

According to the frequency analysis and credibility quantification procedure, radiated noise data is processed as the following.

#### i) Data preprocessing

Radiated noise data preprocessing includes data interception and normalization. Figure 7 reveals the simulation data and reference data in time domain.

#### ii) Welch's spectrum estimation

Using Welch's method to estimate the power spectrum and draw the spectrum graphics. In Figure 8, since the spectrum power between 1 Hz~1000Hz is above 0dB, the band 1Hz~1000Hz is the key band in the following analysis. There emerge seven line spectrums in the Welch spectrum between 1Hz~1000Hz, including 10Hz~54Hz, 82Hz~180Hz, 200 Hz ~240Hz, 320~360HZ, 460Hz~510Hz, 722Hz~785Hz and 850~935Hz.

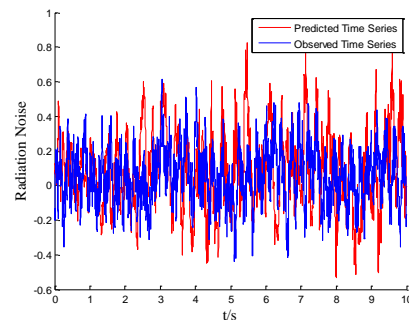
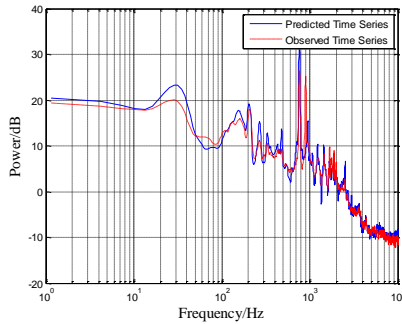


Figure 7. Comparison between predicted and observed data in time domain.



**Figure 8.** Power spectrum estimated by Welch's method.

Analytic hierarchy process (AHP) is used to obtain the weights of every line spectrum. According to the power and band length of each line spectrum, the judgment matrix is defined as:

$$J_D = \begin{bmatrix} 1 & 1/2 & 1 & 1/2 & 1/2 & 2 & 2 \\ 2 & 1 & 1 & 2 & 2 & 3 & 3 \\ 1 & 1 & 1 & 1 & 1 & 2 & 2 \\ 2 & 1/2 & 1 & 1 & 1 & 1 & 2 \\ 2 & 1/2 & 1 & 1 & 1 & 2 & 2 \\ 1 & 1/3 & 1 & 1 & 1/2 & 1 & 1 \\ 1 & 1/3 & 1/2 & 1/2 & 1/2 & 1 & 1 \end{bmatrix}$$

The weight vector of line spectrum is:

$$\omega = [0.122, 0.248, 0.164, 0.150, 0.150, 0.086, 0.080]$$

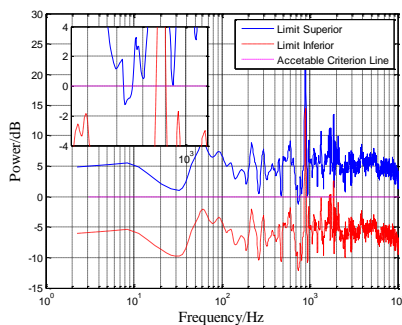
The coincidence index is 1.32 and the ratio is:

$$CR = CI / RI = 0.0158 < 0.1$$

Therefore the weight vector calculated is valid.

**iii) Consistency test**

We can use *F* test to check the behavioral consistency between simulation data and reference data in frequency domain. The consistency analysis result is shown in Figure 9. Under log-log coordinate system, for a frequency point, if the confidence interval contains 0dB, then the radiated noise data passed the consistency analysis on this point.



**Figure 9.** Consistency test result of Welch's power spectrum.

Obviously, there are two frequency bands failed to pass the check, band 736 Hz ~760 Hz and band 866 Hz ~915 Hz.

In figure 8, the power difference of line spectrum 722 Hz ~785Hz excess 5dB. This is the reason for band 736 Hz ~760 Hz failed to pass the consistency

test. Meanwhile, there exists a line spectrum in 858Hz~931Hz for the simulated time series and another in 850Hz~935Hz for the observed time series. The frequency band difference leads to the test failure in band 866 Hz~915 Hz.

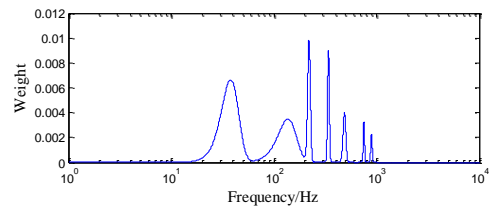
**iv) Credibility quantification**

According to the rule choosing weight function parameters, the parameters of seven line spectrums are defined as Table 1.

**Table 1.** Parameters of Normal Weight Density Function.

Function	Line spectrum	$\mu$	$\sigma$
$F_1$	10Hz~54Hz	38	7.33
$F_2$	95Hz~180Hz	137.5	28.33
$F_3$	200 Hz ~240Hz	220	6.67
$F_4$	320~360HZ	340	6.67
$F_5$	462Hz~512Hz,	487	15.00
$F_6$	722Hz~785Hz	753.5	10.50
$F_7$	850~935Hz	892.5	14.17

Based on the weight of each line spectrum  $\omega$  and parameters of normal weight density function, the weight on the whole frequency band is shown in Figure 10.



**Figure 10.** Weight on the whole frequency band.

Since most part of the line spectrum passed the consistency test, we can calculate the credibility loss

$$\sum_{k \in \Theta} L_k \text{ and the credibility is } 1 - \sum_{k \in \Theta} L_k .$$

The credibility loss  $L_1$  in band 736 Hz ~760Hz is:

$$L_1 = [F_6(760) - F_6(736)] \times 0.086 = 0.059$$

The credibility loss  $L_2$  in band 866 Hz ~915Hz is:

$$L_2 = [F_7(915) - F_7(866)] \times 0.080 = 0.073$$

Then, the final credibility is:

$$C = 1 - L_1 - L_2 = 0.868$$

As a comparison, we use direct transform approach (Zhang J. Y, 2010) to calculate the credibility:

$$C_{dt} = [(736 - 760) + (915 - 866)] / 1000 = 0.927$$

From the Welch's spectrum estimated result and consistency test result, the behavior of simulated time series and observed time series are similar to each other in most parts of the frequency band except two. The credibility quantification result confirms this

conclusion. The credibility calculated by direct transform approach is about 0.927, which is lack of factual basis. To summarize, the credibility quantification approach should focus on key line spectrums and the case study proves the quantification method proposed is effective for time series with complicated spectrum.

## 5 Conclusions

Based on Welch's periodogram and consistency test approach, a novel credibility quantification method using weight density function is proposed. Compared to the direct transform approach, which assume all the point have same effect on the final credibility, the credibility quantification method proposed based on normal weight density function focus on the transform of Welch's analysis result on key frequency band. The credibility quantification of ship radiated noise data proves the transform approach is reasonable. Meanwhile, the credibility quantification process indicates the method provided is effective for periodic time series with complicated spectrum.

Frequency-domain analysis in model validation is a kind of consistency test method based on pattern. Even though there exists uncertainty in simulation model input, frequency-domain approach, including Welch's analysis, Maximum Entropy Spectral Estimation (MESE) et al, can still be utilized to analyze the periodic time series. For the time series with sophisticated spectrum, extraction of line spectrum highly depends on the evaluation expert, which is time-consuming and boring for the analyst. In the future, method will be studied to obtain the accurate information of each line spectrum automatically. Furthermore, a frequency-domain validation and credibility quantification tool will be developed to improve the efficiency of model validation.

## Acknowledgment

This work was supported by National Science Foundation of China (No. 61374164).

## References

Z. G. Chen. *Time Series and Spectrum Analysis*. Science Press, 253-282, 1988.

S. Ferson, W. L. Oberkampf and L. Ginzburg. Model validation and predictive capability for the thermal challenge problem. *Computer Methods in Applied Mechanics & Engineering*. 197(29-32), 2408-2430, 2008. doi: 10.1016/j.cma.2007.07.030.

G. S. Fishman and P. J. Kiviat. The analysis of simulation-generated time series. *Management Science*, 13(7), 525-557, 1967. doi: 10.1287/mnsc.13.7.525.

W. Li, W. Chen, C. Jiang, Z. Z. Lu and Y. Liu. New validation metrics for models with multiple correlated responses. *Reliability Engineering & System Safety*, 127, 1-11, 2014. doi: 10.1016/j.res.2014.02.002.

D. C. Montgomery and R. G. Conard. Comparison of simulation and flight-test data for missile systems. *Simulation*, 34(2), 63-72, 1980. doi: 10.1177/003754978003400206.

D. C. Montgomery and L. Greene. Methods for validating computer simulation models of missile systems. *Journal of Spacecraft and Rockets*, 20(3), 272-278, 1983. doi: 10.2514/3.25592.

J. Mullins, Y. Ling, S. Mahadevan, L. Sun and A. Strachan. Separation of aleatory and epistemic uncertainty in probabilistic model validation. *Reliability Engineering & System Safety*, 147, 49-59, 2015. doi: 10.1016/j.res.2015.10.003.

W. L. Oberkampf and T. G. Trucano. Verification and validation benchmarks. *Nuclear Engineering and Design*, 238(3), 716-743, 2008. doi: 10.1016/j.nucengdes.2007.02.032.

C. J. Roy and W. L. Oberkampf. A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. *Computer Methods in Applied Mechanics and Engineering*, 200(25-28), 2131-2144, 2011. doi: 10.1016/j.cma.2011.03.016.

R. G. Sargent. Verification and validation of simulation models. *Journal of Simulation*, 7(1), 12-24, 2013. doi: 10.1057/jos.2012.20.

S. Sankararaman and S. Mahadevan. Model validation under epistemic uncertainty. *Reliability Engineering & System Safety*, 96(9), 1232-1241, 2011. doi: 10.1016/j.res.2010.07.014.

Z. C. Wang. Research on simulation theory. *Journal of System Simulation*. 12(6), 604-608, 2000. doi: 10.3969/j.issn.1004-731X.2000.06.007.

P. D. Welch. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2), 17-20, 1967. doi: 10.1109/tau.1967.1161901.

J. Y. Zhang. *Validation methods and assistant tools based on coherence of data*. Master Thesis of Harbin Institute of Technology, 32-34, 2010.

Z. Zhang, K. Fang and M. Yang. Method for complex simulation credibility evaluation based on group AHP. *System Engineering and Electronics*, 33(11), 2569-2573, 2011. doi: 10.3969/j.issn.1001-506X.2011.11.42.